I chose the loan prosper file to did the data analysis. This data set is the information of the loan borrowers.

Univariate Exploration :

1. I got a histogram of the borrowerAPR to see the distribution of the borrowerAPR, and most of borrowers' APR are concentrated at 0.15-0.3.
2. From the histogram of the borrowers' stated monthly income. I can see the majority borrowers' income is concentrated at 3000-6000.
3. I also got a histogram for the credit score, most of the credit score is concentrated at 650-800.
4. From the histogram of the lender yield, we can see the lender yield is concentrated at 0.06-0.33.
5. From the histogram of terms, the majorities are about 40 or 60.
6. From the investors histogram, I can see that most of the investors is about 0-100.

Bivariate Exploration :

1. I got the heat map of the stated monthly income and credit score. From the graph, I can see most borrowers' credit score and income is concentrated at credit score between 650-750, stated monthly income between 3000-5000.

2. In order to get the monthly stated income level and borrower APR association. I created a new column income_level. Based on the statistics of the stated monthly income, I categorized the stated monthly income into 4 groups. I made a violin plot for borrower APR for each income level group. The plot showed that the borrower APR distributed very similar among each group, but the high and very high income group have more borrower APR concentrated at 0.1-0.2.

3. In order to get the credit score level and lender yield association. I created a new column called credit score level. From the box plot, I can see that the higher the credit score, the lower of the lender yield.

Multivariate Exploration:

1. I got a scatter plot for borrower APR and lender yield of each credit level and income level. The graph showed that, the the higher of the borrower APR, the higher of the lender yield, and they are linear associated. And the association is not influence by the income or credit level.